DOCUMENT RESUME

ED 423 465 CG 028 679

AUTHOR Lichtenberg, James W.; Hummel, Thomas J.
TITLE The Communication of Probabilistic Information through Test Interpretations.
PUB DATE 1998-08-00
NOTE 32p.; Paper presented as part of a symposium at the American Psychological Association Annual Convention (106th, San Francisco, CA, August 14-18, 1998).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Counseling Psychology; Counselor Training; *Interpersonal Communication; *Predictive Measurement; *Psychological Testing; *Test Interpretation

ABSTRACT

The application of psychological testing is largely an attempt to derive probabilistic statements regarding the likelihood of occurrence of client states, choice outcomes, situational antecedents, and behavioral outcomes. Counseling psychologists share with others the occupational requirement of having to deal with uncertainty in test results and communicate it to others. The use and meaning of probabilistic expressions used in the context of psychological test interpretation is investigated. Specifically, the quantitative meanings of verbal probability expressions used in two different assessment reports are examined with a goal of describing the variability of meanings ascribed to probabilistic terms or phrases used within the reports. Samples were taken from three different training programs (N=66). Results are presented in tables and discussed. Results indicate considerable variability among participants in the meanings they attribute to probabilistic expressions used in the reports and considerable overlap among the words or expressions studied. Differences in the meanings attributed to probabilistic phrases suggest possible differences in the ways counseling psychologists are taught to understand the language of test interpretation reports. (Author/EMK)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 465

Running Heading: PROBABILISTIC INFORMATION

The Communication of Probabilistic Information Through Test Interpretations

James W. Lichtenberg, Ph.D.
University of Kansas

Thomas J. Hummel, Ph.D.
University of Minnesota

Paper presented as a part of a symposium on "Test interpretation in counseling--Recent research and practice" (C. Claiborn, Chair) at the annual convention of the American Psychological Association, August, 1998, San Francisco, CA.

This investigation was supported in part by University of Kansas General Research allocation #3614-XX-0038.

CG028679

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. LICHTENBERG

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

We investigated the use and meaning of probabilistic expressions used in the context of psychological test interpretation. Specifically, we examined the quantitative meanings of verbal probability expressions used in two different assessment reports, with a goal of examining and describing the variability of meanings ascribed to various probabilistic terms or phrases used within the reports. Results indicated considerable variability among participants in the meanings they attribute to probabilistic expressions used in the reports. Although differences were found among the mean probability ratings assigned to the various words/expressions, the results suggest considerable overlap among the words/expressions. Differences in the degrees of variability in ratings across expressions were not significant. Differences in the meanings attributed to the probabilitistic words/expressions were found among the three samples (training programs), suggesting possible "training program differences " in the way that counseling psychologists may be taught to understand the language of test interpretation reports.

The Communication of Probabilistic Information Through Test Interpretations

Introduction

Counseling psychologists, like many other professionals (e.g., teachers, school psychologists, physicians, meteorologists, political scientists), share the occupational requirement of having to deal with uncertainty and communicating it to others. The physician predicts a rash will "probably" go away; a political scientist predicts the Democratic candidate is "almost certain" to win; the meteorologist predicts that rain is "likely" tonight. The counseling psychologist predicts that the client is likely to find a particular occupation or course of study satisfying, or that the child is unlikely to succeed in a regular classroom setting, or that the client may attempt suicide, or that the client probably was abused as a child, or that the client almost certainly is a child molester, or that it is possible that the client will become violent, or s/he states that the client occasionally has flashbacks, or that individuals with this profile are fairly common.

The application of psychological testing is in large an attempt to derive probabilistic statements regarding the likelihood of occurrence of client states, choice outcomes, situational antecedents, and behavioral outcomes. Grounded in psychometric theory, psychological tests are an attempt to quantify these probabilities, and directly or indirectly, psychological test interpretation--whether done clinically or mechanically (Goldman, 1973)--is an attempt to translate and express those probabilities into words rather than numbers.

Test interpretations, written or oral, may be made to clients, sanctioners of services (e.g., parents, the courts, employers), fellow professionals, or others with a legitimate need and right to know. How the recipients of such interpretations translate these qualitative descriptions of behavioral probabilities into numerical estimates of attributes is unclear, although there is

considerable evidence drawn from literature outside of counseling psychology to suggest that it would be unwise to assume that the message sent carries the same meaning as the message received (Budescu & Wallsten, 1985; Reagan, Mosteller & Youtz, 1989; Sutherland, et al., 1991). Indeed, despite a common formal training in psychometrics and in the use of specific tests, the evidence would suggest that counselors as communicators of test interpretations themselves are unlikely to share common (quantitative) meanings for the probabilistic expressions they use in test interpretations. Although numerous studies of the subjective and communicative meaning of probabilistic phrases have been conducted (e.g., Bass, Cascio, & O'Connor, 1974; Beyth-Marom, 1982; Brun & Teigen, 1992; Budescu & Wallsten, 1985; Clarke, Ruffin, Hill, & Beamen, 1992; Foley, 1959; Johnson, 1973; Lichtenstein & Newman, 1967; Ness, 1995; Simpson, 1944, 1963; Wallsten, Budescu, Rapoport, Zwick & Forsyth, 1986), none appears to have been conducted within the context of psychological test interpretation.

Most of the empirical studies conducted to date on the meanings of probabilistic words and expressions have involved having individuals assign numerical equivalents to various probabilistic phrases. The results of this research have been consistent: When statements such as "unlikely," "probably," "may," "often," etc. are used, there has been significant variability in the recipients' understanding of the probabilities associated with those terms (between-subject variability) and considerable overlap among the terms. Wallsten, Budescu, Rapoport, Zwick and Forsyth (1986) noted that this finding of significant between-subject variability has been consistent across a number of studies (Bass, Cascio, & O'Connor, 1974 ; Beyth-Marom, 1982; Budescu & Wallsten, 1985; Foley, 1959; Johnson, 1973; Lichtenstein & Newman, 1967; Simpson, 1944, 1963; Sutherland, Lockwood, Tritchler, Sem, Brooks, & Till, 1991). They also noted that although within-subject variability in the assignment of numbers to probabilistic statements was considerably less than that between subjects, it was not minor—a finding that has been consistent across a number of these same studies (Bass, Cascio, & O'Connor, 1974; Beyth-Marom, 1982; Budescu & Wallsten, 1985; Johnson, 1973).

Budescu and Wallsten (1985) investigated college students' probability estimates and rank ordering of a variety of probability phrases (e.g., rarely, seldom, usually not, unlikely, frequently, probable, often, usually, likely) and found that although individuals have relatively stable rank orderings of these phrases, different individuals have different rank orders. That is, although for individuals such words carry a consistent ordered or ranked meaning, between individuals the words may communicate very different probabilities. Clarke, Ruffin, Hill and Beamen (1992) found high levels of within-subject and between subject variability in the use of verbal expressions probability, and they concluded that such expressions lead to very imprecise communication.

In another study, Brun and Teigen (1988) investigated the communication value of verbal probabilistic phrases ("likely," "possibly," "probably," "perhaps"). Finding differences between groups of people in the probabilistic meaning assigned to these words, they concluded that people often misunderstand the intended statistical meaning of the words and phrases. Their results also suggested that the context within which the probabilistic phrases are used contributes importantly to the variability in meaning attributed to the words. This finding replicated in part those of Beyth-Marom (1982) and of Sutherland, et al. (1991).

In a study of the communication of probabilistic information to cancer patients, Sutherland, et al. (1991) found there was no consensus about the numerical meanings of a given word, and they concluded that there appears to be a great deal of "noise" in the communication between patients and health professionals. Their results demonstrated that health care professionals cannot assume that patients, as a group, share the same numerical interpretations of probabilistic words and phrases. They also cited another study (Sutherland, Lockwood, & Till, 1990) in which they found "a disturbingly large proportion" of patients that had difficulty interpreting the probabilistic statement appearing on a treatment consent form. Considering the evidence that also shows there to be poor agreement among health care professionals about the meaning of probabilistic statements (e.g., Kenney, 1981; Kong, Barnett, Mosteller, & Youtz, 1986; Toogood, 1980), Sutherland et al. (1991) expressed concern that patients may be sent mixed-messages by the health care professionals who interact with them.

We believe that the findings and conclusions of Sutherland and his colleagues have particular relevance for counseling psychologists and their work with clients. Although study of the clinical meaning assigned to probabilistic words and phrases used within psychological test interpretations has yet to be investigated, their research (as well as other studies) suggests the potential for significant misunderstanding between counselors and clients and between counselors and colleagues.

The issue of the meaning of probabilistic phrases is also an issue with respect to scale or instrument construction and the subjective probabilistic meaning assigned to points on rating scales (e.g., Likert scales). Ness (1995) for example found not only that individuals differed in the probabilistic meanings associated with rating scale anchor points, but also that the meanings of scale ratings depended on the scaling method used (rank ordering the words/phrases, estimated percentages associates with the words/phrases, or assigning the words to successive points/intervals along a 7-point scales). He found that the ordinal position (rank) of identical terms would vary depending on the rating method used. As already noted, the scaling of terms and phrases that constitute the probability dimension (e.g., "unlikely," "possible," "very likely) also have been scaled and studied by other researchers (e.g., Bass, Cascio, & O'Connor, 1974; Beyth-Marom, 1982; Budescu & Wallsten, 1985; Clarke, et al., 1992; Lichtenstein & Newman, 1967; Reagan, Mosteller, & Youtz, 1989; Simpson, 1944, 1963; Sutherland, et. al., 1990, 1991), and the results of these studies suggest that identical self-reports provided by respondents to rating scales (e.g., clients) are likely to vary considerably in their meaning to the respondents and to the reviewers of those ratings.

It should be clear that when counselors use terms such as "probable" or "possible" with respect to the meaning of test scores, they intend to convey a meaning or interpretation that implies a certain degree of probability. In a reciprocal fashion, the receiver of the expression interprets or understands a certain degree of probability associated with the words used by the counselor. Confusion, or at least miscommunication, is likely to result if the meaning attached to a probability expression by a counselor is significantly different from the meaning assumed by the recipient of

the expression. If, for example, "probable" means "about 50% of the time" to the counselor and "about 80% of the time" to a child's parent, their individual understandings and decisions regarding the child might be quite different, and the differences in the course of action taken with regard to the child may be significant.

Given the apparent vagueness of probability terms (which would appear to be likely within the context of test interpretation), it is reasonable to ask why actual numbers, percentages, and numerical estimates would not be a preferred means of communicating uncertainties. Wallsten, Budescu, Rapoport, Zwick and Forsyth (1986) suggest that on purely anecdotal grounds, the imprecision of nonnumerical terms seems preferred to the precision of probability numbers for at least two reasons. First, test interpretations (which derive from test scores which in turn are a function of the less than perfect reliability and validity of the measures) are necessarily imprecise, and therefore it would be misleading to represent them with "numerical precision." In this regard, they quote a committee of the U.S. National Research Council. Writing with regard to formal risk assessment, the committee commented that numbers denote authority and a precise understanding of relations, and that there is an

> important responsibility **not** to use numbers, which convey the impression of precision, when the understanding of relationship is indeed less secure. Thus, while quantitative risk assessment facilitates comparison, such comparison may be illusory or misleading if the use of precise numbers is unjustified. (National Research Council Governing Board Committee on the Assessment of Risk, 1981, p. 15; emphasis added)

The second reason suggested by Wallsten et al. for communicating with nonnumerical terms rather than with probability numbers is that most people feel they better understand words than numbers and, therefore, that interpretations are better conveyed verbally than numerically. In this regard,they cite Zimmer (1983) who commented that verbal expressions of uncertainty were available long before the development of mathematical probability concepts, noting that it was not until the 17th century that probability concepts were formally developed while expressions for

different degrees of uncertainty existed in many languages long before then. Zimmer further has suggested that people process uncertainty in a verbal rather than a numerical manner and that judgments are revised in light of new information according to linguistic, rather than numerical, principles.

Despite our best efforts to avoid ambiguity and to enhance the clarity of interpretations, the use of language to communicate probabilities results in a significant likelihood that what we share about our clients with those clients, with their parents or guardians, with professional colleagues, and with the courts, will be understood in the ways other than we intend. The implications of such language imprecision can be significant for all concerned, as such information is used to make important life decisions concerning the person tested. For example, a client may be hospitalized (or released from hospitalization) on the strength of the interpretation provided. A student may be advanced or held back in school based on probability estimates provided regarding the child's likelihood of success in the next grade. Parents may seek or terminate special education services based on their belief regarding the likely benefit of such services--a belief shaped by the interpretation of their child's psychological testing.

In light of the above, it was the purpose of this study to investigate the use and meaning of probabilistic expressions used in the context of a psychological test interpretation. Specifically, the study examined the quantitative meanings of verbal probability expressions used in two different assessment reports, with a goal of examining and describing the variability of meanings ascribed to various probabilistic terms or phrases used within the reports.

Method

Participants

Participants were 66 graduate students from three different APA accredited counseling psychology programs (University of Kansas, N= 15; University of Minnesota, N= 24; University of Southern California, N= 27). All participated in this study as a part of a class on psychological testing. In each instance, the course was an initial testing course offered to students in their

respective programs. Demographic data (gender breakdown, age, racial/ethnic group membership) were not collected on the participants in order to assure their anonymity as participants and students and so are not available on our sample.

Materials

Excepts (approximately printed pages) from examples of MMPI and MMPI-2 interpretive reports (The MMPI Report: National Computer Systems [NCS]) were used as stimulus materials. The sample reports are part of the promotional materials for the MMPI/MMPI-2 scoring and interpretative services offered by NCS. Each excerpt contained numerous examples of the type of probabilistic language provided in these reports, and within the reports, the various probabilistic words or phrases on which we wanted to participants to focus were highlighted.[1]

A rating form on which participants indicated numerical probability estimates for various probabilistic expressions highlighted in the reports was developed for the study. On the form, the probabilistic expressions (taken from the interpretive report) were reproduced, along with the corresponding line numbers for the expressions in the report. Accompanying each expression was a rating scale (0-100% in 5 point increments) on which participants indicated their estimate of the numerical value for the various verbal statements/expressions in the report.

Procedures

Materials were distributed to students in three separate introductory testing/assessment classes. Students were provided with the excerpted reports (MMPI and MMPI-2) and the corresponding response sheets. The participants were instructed to read each report and to then go back through the report and mark on the corresponding response sheet their estimate of the numerical values corresponding to each of the highlighted probability words or phrases.

Analysis

Analyses were conducted separately for the MMPI and MMPI-2 reports. In the occasional instance in which the rating for an expression was missing for a participant, the mean of the group (i.e., the participant's academic program) for that expression was used as the participant's rating.

Prior inspection of the words/phrases used in the report suggested that the expressions represented two different "linguistic sets" -- (a) expressions representing likelihood or frequency estimates of some event (e.g., "probably," "may," "often") and (b) expressions representing the degree or quantity of some variable (e.g., "some," "somewhat," "rather"). On the possibility that participants might differ in their use of these two sets of expressions, the two sets of words/phrases were grouped separately based on a rational analysis of the reports, and each group was analyzed separately.

For each report (MMPI, MMPI-2) and each linguistic group or set of expressions (probability/frequency, degree/quantity), analyses were conducted for (a) differences in mean ratings of the expressions, and (b) differences in the variability in the ratings of the expressions. Differences among the three schools in their mean ratings of the expressions also were investigated.

Results

MMPI

Table 1 summarizes the mean, standard deviation, range, minimum, and maximum for each rated expression in the MMPI report. Across all 32 expressions, mean expression ratings ranged from 47.27 to 80.23, suggesting considerable variability in the ratings across the expressions, and rating ranges varied from a low of 60 to a high of 90, suggesting considerable variability in the participants' ratings of individual expressions. (Expressions preceded by an asterisk [*] are degree/quantity expressions; all others are probability/frequency expressions.)

Insert Table 1 about here

Mean ratings: Expression set 1 (probability/frequency expressions, n= 28). A MANOVA revealed a significant difference among the mean ratings of the expressions, $F(27, 39) = 8.20$, $p<.001$. In light of the number of possible between-expression contrasts that could be conducted, we decided to examine expression differences by ranking the 28 expressions from high (80.23) to

low (48.48) in terms of their mean ratings, and to compare each with the expression having the lowest mean rating. The ratings of all but seven of the expressions were found to differ significantly from the lowest. We also ran contrasts between each ranked expression with the expression ranked immediately subsequent to it. Among these contrasts, only two were significant.

Mean ratings: Expression set 2 (degree/quantity expressions, n= 4). A MANOVA revealed a significant difference among the mean ratings of the expressions, $F(3, 63) = 23.48$, $p<.001$. As in the previous analysis, we examined expression differences by ranking the four expressions from high (63.56) to low (47.27) in terms of their mean ratings, and compared each with the expression having the lowest mean rating. The rating of each of the three expressions ranked above the lowest ranked expression was found to differ significantly from the lowest. We also ran contrasts between each ranked expression with the expression ranked immediately subsequent to it. Among these contrasts, the highest ranked expression differed significantly from the #2 ranked expression, and the #3 ranked expression differed from the #4 ranked expression. The #2 and #3 expressions did not differ significantly in their mean ratings.

Rating variability: Expression set #1. In order to analyze differences in the variability in participants' ratings of individual probability/frequency expressions, participants were randomly sorted into eight groups--six groups of 8 and two groups of 9. Following procedures suggested by Kirk (1982), for each of the eight groups (that now could be treated conceptually as an "individual") the within-group variance in rating for each of the expressions was computed and the natural log of that variance was taken and averaged across the eight groups. This average or mean for each expression was the variable on which differences in expression variability were examined (see Table 1).

Because the number of "individuals" (i.e., groups, n=8) now was fewer than the number of variables (n=28), a multivariate analysis could not be conducted. However, we did run contrasts among expressions to examine possible differences in the variability in the rating of the expressions. As before, in light of the number of possible expressions contrasts that could be

conducted, we examined expression differences by ranking the 28 expressions from high (5.89) to low (4.72) in terms of their mean variability ratings, and compared each with the expression having the lowest mean variability rating. Although one contrast was statistically significant, in light of the number of contrasts conducted, the results of the analysis gave us no reason to believe there to be differences in the variability of the ratings of the expressions. Still using the computed mean variability ratings, we also ran contrasts between each ranked expression with the expression ranked immediately subsequent to it. In this instance, no difference were found, and we again concluded there to be no differences in the variability of the ratings of the probability/frequency expressions.

Rating variability: Expression set #2. The same sorting of participants into eight groups was used in this analysis (see above). This time, however, the four degree/quantity expressions (expression set #2) were analyzed (see Table 1). Because the number of "individuals" (i.e., groups, n=8) now was greater than the number of variables/expressions (n=4), a multivariate analysis could be conducted. The results of this analysis suggested no differences in the variability of the ratings among the four expressions, $F(3, 5)=.176$, $p>.05$. We nevertheless did run contrasts among expressions to examine possible difference in the variability in the rating of the expressions. We initially examined expression differences by ranking the four expressions from high (5.41) to low (5.18) in terms of their mean variability ratings, and compared each with the expression having the lowest mean variability rating. None of the contrasts was significant. We also ran contrasts between each ranked expression with the expression ranked immediately subsequent to it, and again no significant differences were found.

Comparisons among schools. We compared the mean expression ratings among the three schools represented in our sample (University of Kansas, University of Minnesota, University of Southern California). As above, separate analyses were conducted on expression set #1 (probability/frequency) and expression set #2 (degree/quantity).

Mean ratings: Expression set #1. Results of the MANOVA on the probability/frequency expressions indicated a significant difference among the three schools, $F(56, 72)=1.59$, $p<05$.

Significant difference among the three schools were found on 3 of the 28 expressions. Although contrasts between the schools were conducted on the expressions, we do not consider them informative at this time.

Mean ratings: Expression set #2. Result of the MANOVA on the degree/quantity expressions indicated a significant difference among the three schools, $F(8, 120)=2.08$, $p<.05$. A significant difference among the schools was found on one of the four expressions, however, as with the previous contrasts between schools, we did not consider the between school differences on this expression to be informative.

MMPI-2

Our analyses of the ratings of the probabilistic words/phrases in the MMPI-2 report parallel those of the previous MMPI report. Table 2 summarizes the mean, standard deviation, range and minimum and maximum for each rated expression in the MMPI-2 report. Across all 30 expressions, mean expression ratings ranged from 47.95 to 74.77, suggesting considerable variability in the ratings across the expressions, and rating ranges varied from a low of 60 to a high of 85, suggesting considerable variability in the participants' ratings of individual expressions. (As with Table 1, expressions preceded by an asterisk [*] are degree/quantity expressions; all others are probability/frequency expressions.)

Insert Table 2 about here

Mean ratings: Expression set 1 (probability/frequency expressions, n= 25). A MANOVA revealed a significant difference among the mean ratings of the expressions, $F(24, 42) = 7.96$, $p<.001$. As before, in light of the number of possible expressions contrasts that could be conducted, we examined expression differences by ranking the 25 expressions from high (70.76) to low (47.95) in terms of their mean ratings, and compared each with the expression having the lowest mean rating. Fourteen of the expression ratings were found to differ significantly from the lowest expression rating. We also ran contrasts between each ranked expression with the expression ranked immediately subsequent to it. Among these contrasts, none was significant.

Mean ratings: Expression set 2 (degree/quantity expressions, n= 5). A MANOVA revealed a significant difference among the mean ratings of the expressions, $F(4, 62) = 30,28$, $p<.001$. As in the previous analysis, we examined expression differences by ranking the five expressions from high (74.77) to low (50.68) in terms of their mean ratings, and compared each with the expression having the lowest mean rating. The ratings of each of the first three expressions ranked above the lowest ranked expression were found to differ significantly from the lowest rating; the rating of the expression ranked immediately above the lowest ranked expression did not differ from the lowest rated expression. We also ran contrasts between each ranked expression with the expression ranked immediately subsequent to it. Among these contrasts, the highest ranked expression differed significantly from the #2 ranked expression, and the #2 ranked expression differed from the #3 ranked expression. The rating of the #3 ranked expression did not differ from that of the #4 ranked expression, and the rating of the #4 ranked expression did not differ from that of the #5 ranked expression.

Rating variability: Expression set #1. As with our analysis of the MMPI ratings, in order to analyze for differences in the variability in participants' ratings of individual probability/frequency expressions in the MMPI-2 report, participants were randomly sorted into eight groups--six groups of 8 and two groups of 9. (Note: This was a separate random sorting from that used in our analysis of the MMPI ratings.) As before, for each of the eight groups, the within-group variance in ratings for each of the expressions was computed, and the log of that variance was taken and averaged across the eight groups. This average or mean for each expression was the variable on which differences in expression variability were examined (see Table 2).

Again, the number of "individuals" (i.e., groups, n=8) was fewer than the number of variables (n=25), and so a multivariate analysis could not be conducted. As before, however, we did run contrasts among expressions to examine possible differences in the variability in the rating of the expressions. We then examined expression differences by ranking the 28 expressions from high (6.48) to low (4.81) in terms of their mean variability ratings, and compared each with the

expression having the lowest mean variability rating. Seven of the 24 contrasts were statistically significant, providing some evidence for differences in the variability in the expression ratings across the MMPI-2 report. Still using the computed mean variability ratings, we also ran contrasts between each ranked expression with the expression ranked immediately subsequent to it. In this instance, only two of the contrasts were statistically significant.

Rating variability: Expression set #2. The same sorting of participants into eight groups was used in our analysis of the five MMPI-2 degree/quantity expressions (expression set #2) (see Table 2). Because the number of "individuals" (i.e., groups, n=8) was greater than the number of variables/expressions (n=5), a multivariate analysis could be conducted. The results of this analysis suggested no differences in the variability of the ratings among the four expressions, $F(4, 4)=1.65$, $p>.05$. We nevertheless did run contrasts among expressions to examine possible differences in the variability in the rating of the expressions. We first examined expression differences by ranking the five expressions from high (5.44) to low (4.65) in terms of their mean variability ratings, and compared each with the expression having the lowest mean variability rating. The ratings for the expressions ranked 1, 3 and 4 each differed significantly from the rating of the lowest ranked expression; the rating for the #2 ranked expression did not differ significantly from that of the #5 ranked expression. We also ran contrasts between the rating of each ranked expression with the rating of the expression ranked immediately subsequent to it. In this instance, the only statistically significant contrast was between the #4 and #5 ranked expressions.

Comparisons among schools. As with the MMPI expression ratings, we compared the mean MMPI-2 expression ratings among the three schools represented in our sample. As before, separate analyses were conducted on the expression set #1 (probability/frequency) and expression set #2 (degree/quantity).

Mean ratings: Expression set #1. Results of the MANOVA on the probability/frequency expressions indicated no significant differences among the three schools, $F(50, 78)=1.09$, $p>05$, and between school contrasts on the expressions were not considered.

Mean ratings: Expression set #2. Result of the MANOVA on the degree/quantity expressions indicated no significant differences among the three schools, $F(10, 118)=1.65$, $p>.05$, and so between school contrasts on the expressions were not considered.

Discussion

It is important to be clear that the information communicated via test interpretations is probabilistic in part because of the measurement and prediction error issues associated with test reliability and validity. Psychological measurements (i.e., test scores) are not completely consistent. If a client is measured twice on the same measure, even on the same day, those measurements/scores are likely to differ. Similarly, predictions based are test scores are not without error.

Although the information reflected in a test score is necessarily probabilistic, at the same time, most test scores and the predictions derived from them are not completely random, and methods of studying, defining and estimating the consistency or inconsistency of test scores form the central focus of research and theory dealing with the reliability of test scores. Methods of studying test scores and estimating their relationship to others measured behaviors is the focus of research dealing with the criterion-related validity of the test scores.

The relationship between a predictor (or set of predictors) and the criterion is rarely perfect; inevitably there is error in prediction. Likewise, our interpretations of test scores--the verbal comments we provide on the association between a test score and its behavioral correlate(s) are necessarily probabilistic--combining (often in uncertain ways) both the error in measurement and error in prediction. Although these sources of error are the basis for needing to offer interpretations that are probabilistic, at issue in this study was the way in which these probabilistic statements are interpreted by the counselors (or clients) and the possible compounding of the "inaccuracy" of the information that tests provide to others.

The practical importance of consistency in test scores and the predictions we derive from them is a direct result of the fact that tests are used to make important decisions about people. The

same can be said about test interpretations. In counseling psychology, as in a number of other areas of applied psychology, we make heavy use of judgments of probability or frequency and of judgments of degree or amount. In our clinical interpretations of tests and in our clinical decision making which is based (in part) on tests, we are called upon to describe or estimate how often or to what degree the behavioral statements apply to the client being evaluated.

Although study of the clinical meaning assigned to probabilistic words and phrases used within test interpretations had not previously been investigated, the research reviewed in the introduction to this paper (as well as other studies) suggested the potential for significant misunderstanding between therapist and client and between therapist and colleague. The results of our study would seem to support this notion.

Using excerpts from actual interpretative reports of the MMPI and MMPI-2, we found significant variability in the way in which counselors-in-training understood or interpreted the probabilistic language in the reports. Far from communicating a common interpretation, these "standardized" reports resulted in rather strikingly large ranges in the interpretation of words such as "probably," "likely," etc. That is to say, the readers of the reports did not understand the meaning and implications of the interpretations in the same way-- results that replicate the findings of Beyth-Marom (1982), Lichtenstein and Newman (1967), Simpson (1944, 1963), and numerous others.

To the extent that different understandings of the same interpretive language is a problem for counselors, clients and others who may be the recipients of the interpretation, Table 1 illustrates the size of that problem. Inspecting the minimum and maximum ratings reveals ratings on the same individual probability expression as different as 10% and 100% -- a range of 90 percentage points (see expression line # 69B and #75). If one were to assume that the ratings for each probability expression to be normally distributed about the mean for the expression, and if one were to use (for purposes of this example) the average standard deviation across the 32 expressions in Table 1 (SD_{ave}=15.5), one would know that 16% of the recipients of the interpretive

expression would have rated the expression more than 15.5 points lower and 16% would have rated the expression more than 15.5 points higher than the average for a given expression.

One way to think about the utility of probabilistic language is to consider the extent to which different words and phrases communicate different meanings. To the extent that different probabilistic words or phrases are understood to mean the same thing, they fail to provide the receiver of the communication with discriminating and useful information. The results of our study suggest that although a few significant differences were found between the ratings (attributed meanings) of certain probabilistic words or phrases in the reports, the differentiations made among those words/phrases were relatively few. That is to say, not only did the words not mean the same to everyone (as evidenced by the variance in the ratings made for individual items), but the degree of discrimination among the words (i.e., variability in the means) was fairly limited. These results appear to parallel those of others (e.g., Bass, Cascio, & O'Connor, 1974; Borges & Sawyers, 1974; Brun & Teigen, 1988; Clarke, Ruffin, Hill, & Beamen, 1992).

Another way to think about the utility of interpretive language and test interpretation communications is to consider the variance in the meaning attributed to the words and phrases used in interpretive reports. Words or phrases with greater variance generally would be less useful, as their variance would reflect a lack of consistency or precision in their interpretation. Although our analyses suggested few differences among the words and phrases in terms of the degree of variability in their ratings (interpretations), such a finding does not mitigate our previous finding of considerable variability in the meaning attributed to the language in the reports. Rather, it simply suggests that none of the words or phrases is necessarily more (or less) clear in its interpretation.

It was interesting, but a bit disconcerting, to find that there appeared to be some systematic differences in the interpretation of the MMPI report that could be attributed to the school or program in which the participants were enrolled. This finding suggests that the meaning given to a report might depended on the program from which a counseling psychologist is graduated. The reason for these differences is not clear, but the difference raise the uncomfortable possibility that meaning given to interpretive results of tests and the clinical decisions based upon those results

may be more a function of the counselor's academic training program than of the test results themselves.

Clearly, verbal probability expressions are a poor tool to convey one's test interpretations--be they descriptive, genetic, predictive, or evaluative (see Goldman, 1973). A decision maker (whether that be the therapist, client, parent, teacher, etc.) receiving a test interpretation may understand the event probability very differently from the way the person proffering the interpretation intended and may base an important decision on an erroneous understanding of the results of testing.

Beyth-Marom (1982) suggested that one might be tempted to discount disagreements in understanding of probabilistic language found in earlier studies on the grounds that probability expressions normally are used in specific contexts which tend to decrease their range of interpretation (e.g., see Brun & Teigen, 1988). However, in her study and others (e.g., Sutherland, et al., 1991), disagreements in the interpretation of verbal probability expressions (as opposed to words or expressions presented to individuals in isolation) were actually higher when assessed "in-context." Although in our study we did not contrast participants' ratings of verbal probability expressions within the context of the MMPI and MMPI-2 reports with similar expressions presented "in isolation," the considerable variability in the ratings assigned by participants to words/expressions in the reports suggests the "within-context" understanding of the probabilistic language of the reports to be extremely varied.

To paraphrase Beyth-Marom (1982), the results of her study (and of ours) should convince any prognosticating psychologist to use numerical expressions of probability rather than verbal ones—a point that might be made to test interpretation services such as IPAIT, NCS, Psych Corp., etc. A similar point is made by Kenney (1981) and by Nako and Axelrod (1983) in the area of medical diagnosis and prediction.

Specifically, Kenney (1981) has suggested, based on an informal study of physicians at Massachusetts General Hospital, that when trying to communicate probabilistic information clearly to others, certain terms should be avoided (those identified by a large range and standard

deviation). He also suggested that clinicians, authors, and editors should consider the imprecision of their terms, and if semiquantitative terms must be used for lack of hard data, it might be wise to include (in parentheses) their best estimate of the value or range that they are trying to convey. With regard to the practice of offering interpretations of psychological and educational tests, we believe that Kenney's suggestion may have considerable merit

References

Bass, B. M., Cascio, W. F., & O'Connor, E. J. (1974). Magnitude estimations of expression of frequency and amount. Journal of Applied Psychology, 59, 313-320.

Beyth-Marom, R. (1982). How probable is probable? A numerical translation of verbal probability expressions. Journal of Forecasting, 1, 257-269.

Borges, M. A., & Sawyers, B. K. (1974). Common verbal quantifiers: Usage and interpretation. Journal of Experimental Psychology, 102, 338-340.

Brun, W, & Teigen, K. H. (1988). Verbal probabilities: Ambiguous, context-dependent, or both? Organizational Behavior and Human Decision Processes, 41, 390-404.

Bryant, G. D., & Norman, G. R. (1980). Expressions of probability: Words and numbers. New England Journal of Medicine, 301, 411.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. Organizational Behavior and Human Decision Processes, 36, 391-405.

Clarke, V. A., Ruffin, C. L., Hill, D. J., & Beamen, A. L. (1992). Ratings of orally presented verbal expressions of probability by a heterogeneous sample. Journal of Applied Social Psychology, 22, 638-656.

Foley, B. J. (1959). The expression of certainty. American Journal of Psychology, 72, 614-615.

Goldman, L. (1973). The use of tests in counseling. New York: Appleton-Century-Crofts.

Johnson, E. M. (1973). Encoding of qualitative expressions of uncertainty. (Technical Paper No. 250). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Kenney, R. M. (1981). Between never and always. New England Journal of Medicine, 305, 1097.

Kirk, R. E. (1982) Experimental design for the behavioral sciences (2nd ed.). Monterey, CA: Brooks/Cole.

Kong, A., Barnett, G. O., Mosteller, F., & Youtz, C. (1986). How medical professionals evaluate expressions of probability. New England Journal of Medicine, 315, 740.

Lichtenstein, S., & Newman, J. R. (1967). Empirical scaling of common verbal phrases associated with numerical probabilities. Psychonomic Science, 9, 563-564.

Nako, M. A., & Axelrod, S. (1983). Numbers are better than words: Verbal specifications of frequency have no place in medicine. American Journal of Medicine, 74, 1961.

Ness, M. E. (1995). Ordinal positions and scale values of probability terms as estimated by three methods. Measurement and Evaluation in Counseling and Development, 28, 152-161.

Reagan, R. T., Mosteller, F., & Youtz, C. (1989). Quantitative meanings of verbal probability expressions. Journal of Applied Psychology, 74, 433-442.

Simpson, R. H. (1944). The specific meanings of certain terms indicating differing degree of frequency. The Quarterly Journal of Speech, 30, 328-330.

Simpson, R. H. (1963). Stability in meanings for quantitative terms: A comparison over 20 years. Quarterly Journal of Speech, 49, 146-151.

Sutherland, H. J., Lockwood, G. A., & Till, J. E. (1990). Are we getting informed consent from patients with cancer? Journal of Research in Science and Medicine, 83, 439.

Sutherland, H. J., Lockwood, G. A., Trichler, D. L. , Sem, F., Brooks, L., & Till, J. E. (1991). Communicating probabilistic information to cancer patients: Is there "noise" on the line? Social Science and Medicine, 32, 725-731.

Toogood, J. H. (1980). What do we mean by "usually"? Lancet, 1, 1094.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Journal of Experimental Psychology: General, 115, 348-365.

Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretation of probability and frequency expressions. Journal of Memory and Language, 25, 571-587.

Zimmer, A. C.. (1984). A model for the interpretation of verbal predictions. International Journal of Man-Machine Studies, 20, 121-134.

Footnote

[1] Similar narrative reports on the results of other assessment instruments are available from other test distributors, and interpretive reports on a client's MMPI/MMPI-2 are available from other scoring/interpreting services. We used the NCS interpretative reports for the MMPI and MMPI-2 only as examples, and our results are not intended as comments on either the NCS report or on the MMPI/MMPI-2.

Table 1

Mean, Standard Deviation, Range, Minimum, Maximum and Log Variance[1] for Ratings of MMPI Expressions (N=66)

MMPI Expression							
Line #	Text	Mean	SD	Range	Min	Max	ln (variance)[1]
15	there may be some tendency for...	49.31	17.84	80	15	95	5.59
17	she may be...	48.48	16.71	85	5	90	5.53
25	her complaints are probably...	66.97	15.64	70	25	95	5.26
25B	her complaints are possibly...	52.50	14.71	65	20	85	5.24
26	she is probably...	68.48	14.22	65	30	95	5.07
27	she is likely to...	70.98	13.57	60	35	95	5.06
28	she may be...	52.88	16.08	75	20	95	5.29
30	many individuals...	70.53	14.76	70	20	90	5.09
32	[these symptoms] are common...	70.36	14.69	70	25	95	5.26
41	there is a strong possibility that...	80.23	14.15	75	20	95	4.87
*48	she appears to be somewhat...	47.27	15.91	75	15	90	5.21
48B	she may...	49.09	15.24	85	10	95	5.28
50	unhappiness is likely to be...	66.82	13.09	75	20	95	5.05
*51	she is rather...	63.56	14.56	65	25	90	5.18
*52	she has some...	53.18	17.95	85	15	100	5.33
52B	she is probably...	64.02	14.89	60	30	90	5.14
*53	she may be somewhat...	53.11	16.82	75	15	90	5.41
57	she is tending to be...	62.50	15.55	75	20	95	5.28

61	there is likely to be...	65.53	14.52	80	15	95	5.20
61B	her problems predispose her to...	66.03	16.30	70	25	95	5.39
62	her disorder could...	50.15	18.56	75	5	80	5.75
67	such individuals are often...	68.48	15.27	80	15	95	5.39
67B	such individuals may...	52.88	15.54	85	10	95	5.36
69	these problems might be...	50.44	16.33	75	20	95	5.34
69B	some individuals...	48.86	19.88	90	10	100	5.89
71	her scores suggest great proneness...	78.86	13.83	75	20	95	4.72
75	her symptoms are associated...	63.79	18.02	90	10	100	5.56
82	she is likely to...	65.61	13.83	65	30	95	4.97
83	she tends to...	64.09	15.26	75	20	95	5.24
85	she may not be...	49.62	16.63	75	15	90	5.53
87	such individuals often see...	69.99	12.74	55	40	95	4.85
88	such individuals may respond...	54.11	14.47	75	25	100	5.09

[1] ln refers to the natural log function. Variances are based on N=8 or 9 (see text for explanation of the derivation of these values).

Note: Expression numbers preceded by an asterisk (*) are "degree/quantity" expressions; all others are probability/frequency expressions.

Table 2

Mean, Standard Deviation, Range, Minimum, Maximum and Log Variance[1] for Ratings of MMPI-2 Expressions (N=66)

MMPI-2 Expression							
Line #	Text	Mean	SD	Range	Min	Max	ln (variance)[1]
*23	she appears to be quite...	74.77	11.88	60	35	95	4.65
24	she may be feeling	50.98	14.26	65	20	85	4.89
28	she may do...	49.85	15.76	75	15	90	5.25
32	she may avoid..	49.70	14.57	75	15	90	5.05
38	she seems...	60.30	17.60	85	5	90	5.64
43	she may be...	51.36	15.63	65	20	85	5.41
*44	her responses indicate a rather...	64.02	14.99	70	20	90	5.28
*45	she feels somewhat...	53.18	15.21	70	20	90	5.10
*46	she expresses some...	50.68	17.52	75	15	90	4.44
52	she is probably experiencing...	64.32	16.12	75	15	90	5.56
52B	this is possibly due to...	53.33	16.01	60	20	80	5.26
53	her behavior is likely to...	68.56	15.46	75	20	95	5.24
54	she may be experiencing...	50.98	14.79	75	15	90	5.24
*57	she feels a moderate degree of...	55.15	15.91	75	20	95	5.33
65	such individuals tend to...	64.62	14.89	65	30	95	4.93
66	such individuals might express...	47.95	15.64	75	10	85	5.21
72	there could be some...	48.64	18.33	85	10	95	5.34
74	her style in not likely to...	55.08	26.20	85	5	90	6.48

79	individuals like this are often...	70.76	14.60	70	20	90	5.16
84	individuals like this may...	51.06	13.74	65	20	85	4.99
88	individuals like this may...	52.95	14.99	75	20	95	5.15
88B	acting out is a possibility...	52.27	18.63	85	10	95	5.66
89	she is probably experiencing...	67.27	15.04	70	25	95	5.27
91	she may be...	55.00	16.17	80	15	95	5.41
93	such individuals often...	70.15	17.78	85	10	95	5.26
95	some individuals like this...	50.00	17.28	80	10	90	5.33
100	she may feel...	49.92	13.91	75	20	95	4.81
101	she is not likely to...	51.67	23.03	85	5	90	6.22
106	interventions with her could...	52.54	15.80	80	15	95	5.16
110	she holds feelings that could...	54.99	17.54	85	15	100	5.34

[1] ln refers to the natural log function. Variances are based on N=8 or 9 (see text for explanation of the derivation of these values).

Note: Expression numbers preceded by an asterisk (*) are "degree/quantity" expressions; all others are "probability/frequency" expressions.